

Probabilistic Anomaly Detection Approach for Data-driven Wind Turbine Condition Monitoring

Yuchen Zhang, *Member, IEEE*, Meng Li, Zhao Yang Dong, *Fellow, IEEE*, and Ke Meng, *Member, IEEE*

Abstract—Continuous monitoring of wind turbine (WT) operation can improve the reliability of the wind turbine and lower the operation and maintenance costs. To improve the condition monitoring (CM) and fault detection performance on WTs, this paper proposes an artificial intelligence-based probabilistic anomaly detection approach that can not only provide a deterministic estimation of the WT condition but also evaluate the uncertainties associated with the estimation. An abnormal WT condition is detected based on the evaluated uncertainties, to provide a noise-free incipient fault indication. Compared to the conventional deterministic CM approaches with a residual-based anomaly detection criterion, the proposed probabilistic approach tends to accurately detect the faults earlier, which allows more time for maintenance scheduling to prevent WT component failure. The early fault detection ability of the proposed approach was verified on an operational WT in China.

Index Terms—Condition monitoring, fault detection, probabilistic regression, SCADA, wind turbine.

I. INTRODUCTION

THE wind energy industry has been rapidly developing in recent years due to the large-scale availability of wind resources and pressing environmental issues. Compared to gas/steam/hydro turbines in conventional power plants, wind turbines (WTs) are usually subject to an unpredictable and harsh operating environment, which leads to higher failure rates. Unexpected faults or failures of WT components can impact the turbine availability and increase the operation and maintenance (O&M) costs, thereby reducing the overall power production and the cost of energy.

Condition monitoring (CM) refers to the process of continuously monitoring WT parameters and identifying significant changes in critical parameters that can be early indicators of potential component faults [1]. This has been widely recognized as a good solution to prevent WT failures and lower the O&M costs [2], [3]. Existing CM approaches can be divided

into three categories: model-based, signal-based, and data-driven [4]. In model-based CM approaches, such as observer-based methods [5]–[7], numerical models are built to represent the physical characteristics of the WT, which requires accurate physical system knowledge that is difficult to obtain due to the complicated dynamic interaction among the various components. Signal-based approaches determine the conditions of WT components mainly from vibration [8]–[11], acoustic emission [12]–[15] and electrical voltage/current [16]–[19], etc. A survey of signal processing based WTCM approaches has been provided in [20]. This approach generally requires the installation of additional sensors specifically for CM purposes, which inevitably add costs to O&M. However, the data-driven approach that utilizes the data from supervisory control and data acquisition (SCADA) systems is considered as a cost-effective solution for WTCM and has received extensive research interest [21]–[31]. In general, SCADA systems have been installed in the majority of MW-scale WTs principally for performance monitoring purposes, so the SCADA data is readily available for CM without any system reinforcement for data acquisition, which determines the cost-effectiveness of the data-driven CM approaches.

The SCADA measurements are recorded every 10 minutes. They generally do not contain detailed transient information as provided by the vibration signals, thus conventional signal processing techniques are not applicable to interpret SCADA data [23]. Additionally, SCADA data is collected under practical WT operating circumstances where component faults occur randomly, and it is very expensive to obtain sufficient representative data for all types of faults. For these cases, a supervised classification model was used to directly distinguish normal and faulty WT conditions, as has been attempted in [21], [22], but this tends to be impractical. A better approach for data-driven WTCM is anomaly detection which attempts to develop a normal behavior model (NBM) that learns and describes the normal (i.e. no fault) operating conditions of WT components, and then detect the faults once the online component condition significantly deviates from the NBM estimation. The anomaly detection approaches tend to identify the faults in their beginning stages, which can trigger an earlier warning signal to quickly schedule the maintenance actions. Various anomaly detection techniques have been recently adopted for data-driven WTCM. In [23], a data pre-processing method is proposed to facilitate anomaly detection in wind turbine drive-chain systems. In [24], anomaly detection for generator CM was achieved by constructing an NBM using temperature data and a nonlinear state estimation technique. A co-integration

Manuscript received January 3, 2019; revised March 4, 2019; accepted March 7, 2019. Date of publication June 30, 2019; date of current version May 6, 2019. The work was supported in part by the Australian Research Council (ARC) Discovery Grant (DP170103427/180103217), in part by the Fundamental Research Funds for the Central Universities (No. 2017BSCXB58) and the Postgraduate Research & Practice Innovation Program of Jiangsu Province.

Y. C. Zhang (corresponding author, e-mail: mrhighzhang@gmail.com), Z. Y. Dong, and K. Meng are with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia.

M. Li is with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, Jiangsu, China.

DOI: 10.17775/CSEEJPES.2019.00010

method was developed in [25] to detect WT faults based on the co-integration residuals. Additionally, artificial intelligence (AI) algorithms have shown effectiveness in data-driven WTCM owing to the capability of mining nonlinear data relationships without specialist intervention [26]–[31]. In the literature, the majority use neural networks (NN) to build the NBM and define thresholds on the online prediction residuals to identify potential faults [26]–[29]. Other AI-based anomaly detection approaches for WTCM include an adaptive neuro-fuzzy inference system (ANFIS) [30] and an unsupervised approach using spatiotemporal pattern network and restricted Boltzmann machine [31].

Most of the reported anomaly detection approaches for WTCM adopt deterministic NBMs, which only provide firm estimations of the target parameters without considering the uncertainties associated with the estimation. However, due to the statistical nature of data-mining approaches and the stochastic variation of environmental conditions, uncertainties inevitably exist in the SCADA data, NBMs, and the estimation residuals in practice. Such uncertainties, if not quickly captured and investigated, can have significant impact on the CM reliability and the anomaly detection performance. To mitigate these issues, this paper provides insight into the issue of uncertainties in WTCM and proposes a probabilistic anomaly detection approach that can not only provide deterministic estimation on the target parameters but also evaluate the associated uncertainties. The anomaly detection criterion is devised based on the variation in evaluated uncertainties to provide a noise-free emerging fault indication. Compared to the conventional deterministic approaches with residual-based anomaly detection criterion, the uncertainties quantified by the proposed probabilistic approach serve as a more accurate control variable for anomaly detection. Moreover, the probabilistic approach eliminates the filtering of noise, which tends to detect the faults earlier. Two different probabilistic machine-learning algorithms, i.e. Gaussian Processes (GP) and bootstrap-based ensemble NN (BENN), are attempted in this paper, to construct the NBMs, which demonstrates the generic advantage of probabilistic anomaly detection approaches in WTCM. The proposed approach was demonstrated on a SCADA dataset collected from a realistic WT in China; the testing results had a high level of accuracy and also showed the ability of early fault detection of the probabilistic approach, demonstrating that this would be a good choice of probabilistic machine learning algorithm.

II. PROBLEM DESCRIPTION

This paper proposes an anomaly detection approach based on probability, to highlight the uncertainties involved in WTCM, and aims to provide early detection of potential faults of WT components. This section briefly introduces the WTCM concept, and then presents the general framework of anomaly detection approach. The problems of existing anomaly detection approaches and the proposed solutions are also identified.

A. Anomaly Detection Approach for WTCM

CM refers to the process of monitoring the operating

parameters of a physical system. The abnormal changes in critical parameters serve as indications of upcoming faults in the system. When it is applied to WTs, CM can provide important diagnostic information on the health condition of WT components, such as gearboxes, blades, generators, etc. CM is integrated in the O&M system of a WT. According to the condition of the monitored parameters, maintenance activities should be scheduled and taken to prevent the occurrence of component failure or critical malfunction.

Based on the SCADA data, the CM task is to identify how the changes in operating parameters are related to the status of the WT and its components. One straightforward strategy to achieve this is to label the status at each sampling time (usually every 10 minutes) based on the status/fault code provided by the SCADA system, and then train a supervised classification model to directly represent the relationship between the operating parameters and the different status labels [21]. However, in practice, WTs experience faults randomly and there are numerous types of faults that can occur. Since the available SCADA data is collected on site, it is challenging and time consuming to gather sufficient representative samples for each fault scenario, which limits the application of a supervised classification approach. Alternatively, anomaly detection has been identified as a more practical WTCM approach owing to its minimal reliance on class labels, while still achieving the goal of early fault detection [26]. This approach fits a regression-based NBM to describe the dependency between different operating parameters under a healthy state (i.e. fault free). At the online stage, each incoming new observation is compared with the normal operating condition anticipated by NBM and an alarm is raised if a pattern of significant deviation is detected.

The proposed approach in this paper also follows the general anomaly detection framework for WTCM as shown in Fig. 1. The CM is implemented via offline training and online monitoring. The purpose of offline training is to train the NBM using the selected normal condition data, and online monitoring is a continuous process to detect emerging fault patterns and schedule maintenance actions. The functionality of each block in Fig. 1 is briefly described as follows:

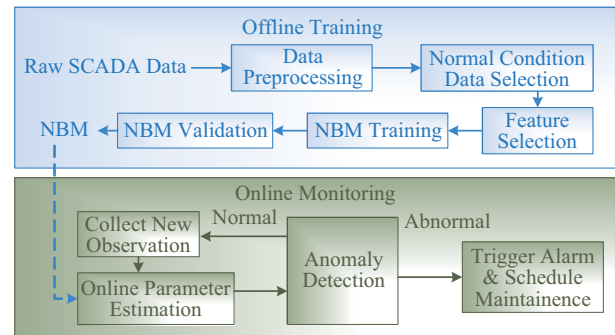


Fig. 1. The anomaly detection framework for WTCM.

1) Data Preprocessing

The raw data provided by a SCADA system is generally not ready for model training and a data preprocessing block is needed as the initial step to ensure data quality. The following

actions are taken to process the raw data: remove invalid observations, deal with missing data, and coordinate time tags.

Additionally, the data needs to be rearranged according to the pre-defined target parameters of an NBM. In general, some operating parameters are critical to WT faults and their abnormal changes can serve as indication of potential faults. Those critical parameters are then the target parameters of an NBM, which means their values will be estimated by the NBM while in operation; as a guide for fault detection. When an abnormally large deviation is observed between the estimated and measured values of a target parameter, warnings or alarms on the corresponding fault type will be raised. In practice, such target parameters are usually defined by the relevant experts based on information such as the WT physical operation mechanism and the risk of faults.

2) Normal Condition Data Selection

This block is to select normal condition observations from the available data. In this paper, the normal condition observations are selected based on the status code provided by SCADA. These status codes can indicate the operation modes and the fault alarms of the WT. To ensure the selected data is from the normal operation period, the observations with status codes “maintenance shutdown”, “fault”, and/or “forced shutdown” are avoided, and on top of that, the observations with future fault alarms issued within t_{ahead} hours are removed. In doing so, the observations at WT downtime and/or with incipient fault patterns are eliminated in the selected normal condition data. The evaluation of t_{ahead} will be further elaborated in Section V.

3) Feature Selection

As the measurements of numerous operating parameters are available in a SCADA system, the feature selection is to select the critical parameters that can best describe each target parameter; this effectively reduces the data dimension for machine learning while achieving low fitting error. The feature selection algorithm used in this paper is RELIEFF which is a statistical learning process to rank the available parameters on their distance-based relevance to the target parameter. Details for more information about RELIEFF can be found in [32] and [33].

4) NBM Training

The NBM is generally a collection of sub-models, each of which describes the normal condition of a target parameter. Considering the complexity of an NBM, the components and target parameters to monitor should be carefully selected based on the failure modes, statistical failure frequency and the average downtime caused by the failures. Training a complete NBM involves the training of individual sub-models which can be processed either in sequence or in parallel depending on the complexity of the training algorithm and the power of the computer to run it. The individual sub-model should be trained based on the selected normal condition data and the critical parameters.

5) NBM Validation

This block is to validate the performance of NBM on normal condition data. Based on the validation results, a criterion to distinguish normal and abnormal conditions needs to be

defined. In this paper, a k -fold cross validation structure is adopted.

6) Online Parameter Estimation

In this block, for each new observation received at the online stage, the validated NBM is used to estimate the value of each target parameter based on its dependent parameters.

7) Anomaly Detection

This block is to detect the abnormality existing in the new observation. This refers to checking the anomaly detection criterion defined in the NBM validation process and triggering alarms when violated. Appropriate maintenance actions should also be scheduled based on the alarm signal.

B. Concerns on Existing Approaches

Most of the anomaly detection approaches reported in the literature adopt a deterministic regression model as the NBM to monitor the WT condition, and the faults of components are detected based on the prediction residual, i.e. the difference between NBM's prediction and the online measurement [26]–[29]. This deterministic approach detects the faults if the prediction residual exceeds its control limit. In practice, due to the statistical nature of the NBM and the stochastic variation of the environment conditions such as wind speed, wind direction, and ambient temperature, the residuals generated by an NBM that is trained with historical SCADA data are subject to noise in representing the normality/abnormality of unseen observations, i.e. the prediction residual can sometimes be unexpectedly high (or low) on normal (or abnormal) condition observations. To smooth out the noise from the prediction residual, the common approach is to use a moving average filter, which is formulated as follows:

$$z_t = \frac{1}{L} \sum_{l=1}^L e_{t-(l-1)\Delta t} \quad (1)$$

where z_t is the smoothed prediction residual, e is the prediction residual, and L is the length of the moving average window. Although the moving average filter is effective in smoothing out the associated noise, it averages the residual over the past and new observations, which induces a lag in the residual variation trend and may delay the fault detection and postpone the maintenance actions.

With these problems in mind, this paper uses a probabilistic NBM that not only estimates the value of target parameters but also evaluates the uncertainties involved in the estimation. The probabilistic model can quickly evaluate the uncertainties without the need of noise smoothing, meaning the potential faults of the WT components can be detected at an earlier stage.

III. PROBABILISTIC NORMAL BEHAVIOR MODELLING

In the proposed approach, the probabilistic regression model is adopted as the NBM for WTCM. This section elaborates the probabilistic NBM and its associated criterion to distinguish normal and abnormal operating conditions of WT components.

A. Probabilistic Regression Model

Conventional deterministic regression models focus on di-

rectly estimating the target parameter value. Although care is taken to improve the model accuracy, regression errors are inevitable and cannot be eliminated. Depending on the data available, such errors generated by the deterministic models can be hard to predict and are difficult to control. In comparison, probabilistic regression models map the unknown target into a distribution rather than a definite value. The results of the distribution format not only provide the information of the most probable target parameter value, but also describe the uncertainties associated with the data, the model, and the estimation process.

Mathematically, for a target parameter y , the probabilistic regression model makes an estimation as follows:

$$y \sim \mathcal{N}(\hat{y}, \sigma^2) \quad (2)$$

$$\hat{y} = F(\mathbf{x}) \quad (3)$$

$$\sigma^2 = G(\mathbf{x}) \quad (4)$$

where the measured target parameter value y falls in a normal distribution with mean $\hat{y}$ and variance σ^2 . In probabilistic prediction, the distribution mean $\hat{y}$ is generally regarded as the fitted target value and σ^2 represents the associated uncertainties. Both of them are derived from the observed input vector $\mathbf{x}$ via different mapping functions $F(\cdot)$ and $G(\cdot)$. These mapping functions represent the probabilistic regression model and they are obtained through offline training. In this paper, GP and BENN are the two machine learning algorithms attempted to construct the probabilistic NBM, which is detailed in Section IV.

B. Anomaly Detection Criterion

Based on the CM results in distribution format, it needs to be determined how to effectively distinguish normal and abnormal conditions. As previously mentioned, using prediction residual as the control variable as per [26]–[29] requires a moving average filter to smooth out the noise estimations, which may delay the detection of emerging faults. In this paper, the anomaly detection criterion is devised based on the distribution variance σ^2 which evaluates the uncertainty associated with each estimation. The variation in σ^2 serves as a noise-free indication of the abnormal condition. For an unseen observation, if its parameters are associated in a way that significantly deviates from the underlying relationship learnt from normal condition observations, the NBM tends to be less certain (i.e. high σ^2) about its estimation, which indicates an emerging fault in the WT component.

Mathematically, the anomaly detection criterion proposed in this paper is as follows:

$$\text{If } \begin{cases} \sigma^2 \leq \eta \Rightarrow \text{Normal Condition} \\ \sigma^2 > \eta \Rightarrow \text{Abnormal Condition} \Rightarrow \text{Trigger Alarm} \end{cases} \quad (5)$$

where η is the control limit on the variance σ^2 , which is derived from the validation results on normal condition observations.

Suppose the variance σ^2 on normal condition observations follows a gamma distribution:

$$\sigma^2 \sim \Gamma(\alpha, \beta) \quad (6)$$

where α is the shape parameter and β is the scale parameter. Those parameters are fitted by the maximum likelihood estimation method on the validation results. Having obtained α and β for the appropriate gamma distribution, the cumulative distribution function is used to establish the control limit, η ,

$$\eta = C^{-1}(p|\alpha, \beta) = \{\eta : C(\eta|\alpha, \beta) = p\} \quad (7)$$

$$p = C(\eta|\alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\eta t^{\alpha-1} e^{-\frac{t}{\beta}} dt \quad (8)$$

where $C(\cdot)$ is the cumulative distribution function, and $p \in [0, 1)$ is the confidence limit which should be large enough to avoid a false alarm.

IV. PROBABILISTIC LEARNING ALGORITHMS

To generalize the performance of a probabilistic regression model in WTCM, two probabilistic learning algorithms, i.e. GP and BENN, are attempted in this paper. This section respectively provides the details of the two algorithms.

A. Gaussian Processes

GP is a Bayesian, non-parametric, and non-linear machine learning algorithm mainly used to deal with regression problems. In essence, GP is a non-parametric generalization of joint normal distribution for a given set of variables [34], [35],

$$F(\mathbf{X}) = \{f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)\} \sim \mathcal{N}(m(\mathbf{X}), K(\mathbf{X})) \quad (9)$$

where $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ consists of the training observations, $m(\cdot)$ is the mean function that is set to zero after normalization, and $K(\mathbf{X})$ is the covariance matrix as follows [34], [35],

$$K(\mathbf{X}) = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) & \cdots & k(\mathbf{x}_2, \mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & k(\mathbf{x}_N, \mathbf{x}_2) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} \quad (10)$$

where $k(\cdot)$ is a covariance function that represents the correlation between any two observations. This covariance function is the core of GP in the sense that it signifies the underlying association between the observations. Although various covariance functions are available as reported in [35], we use a commonly applied function, squared exponential, which is formulated as follows [35],

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left(-\frac{r^2}{2\sigma_1^2}\right) \quad (11)$$

where σ_f and σ_1 are the two hyperparameters, which are tuned by maximizing the log marginal likelihood [34], [35],

$$\log P(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2} \mathbf{Y}^T \mathbf{K}^{-1} \mathbf{Y} - \frac{1}{2} \log |\mathbf{K}| - \frac{N}{2} \log 2\pi \quad (12)$$

where $\boldsymbol{\theta}$ represents the vector containing the hyperparameters.

With the joint distribution in (9) as an initial value for Bayesian inference, a new observation $\mathbf{x}$ collected at online stage can be incorporated as follows [34], [35],

$$\begin{bmatrix} F(\mathbf{X}) \\ f(\mathbf{x}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{X}) \\ m(\mathbf{x}) \end{bmatrix}, \begin{bmatrix} K(\mathbf{X}) & K(\mathbf{X}, \mathbf{x}) \\ K(\mathbf{x}, \mathbf{X}) & k(\mathbf{x}, \mathbf{x}) \end{bmatrix}\right) \quad (13)$$

where

$$K(\mathbf{x}, \mathbf{X}) = [k(\mathbf{x}, \mathbf{x}_1), k(\mathbf{x}, \mathbf{x}_2), \dots, k(\mathbf{x}, \mathbf{x}_N)] \quad (14)$$

$$K(\mathbf{X}, \mathbf{x}) = [k(\mathbf{x}_1, \mathbf{x}), k(\mathbf{x}_2, \mathbf{x}), \dots, k(\mathbf{x}_N, \mathbf{x})]^T \quad (15)$$

The posterior distribution of the target parameter y is

$$y = f(\mathbf{x}) | F(\mathbf{X}) \sim \mathcal{N}(\hat{y}, \sigma^2) \quad (16)$$

where

$$\hat{y} = m(\mathbf{x}) + K(\mathbf{x}, \mathbf{X})K(\mathbf{X})^{-1}(F(\mathbf{X}) - m(\mathbf{X})) \quad (17)$$

$$\sigma^2 = \sqrt{k(\mathbf{x}, \mathbf{x}) - K(\mathbf{x}, \mathbf{X})K(\mathbf{X})^{-1}K(\mathbf{X}, \mathbf{x})} \quad (18)$$

The theory of GP and its covariance functions are well presented in [34] and [35].

B. Bootstrap-based Ensemble NN

In machine learning, ensemble learning refers to a learning paradigm where a collection of single learning units is trained to solve the same problem [36]. Under this paradigm, the individual units can compensate each other, and their combined estimation tends to be more accurate and robust compared to a single learning unit. The key to the success of ensemble learning is the learning diversity among individual units, and such diversity, if numerically identified, can help when evaluating regression uncertainties and the construction of a probabilistic regression model. Following such logic, BENN has been proposed in [37] as a probabilistic wind power prediction method, which is adopted as the second probabilistic learning algorithm in this paper.

The offline training and online application of a BENN is shown in Fig. 2. The BENN consists of two sub-ensembles of NNs. The primary sub-ensemble is used for deterministic estimation on the target parameter while the secondary one is devised to evaluate the associated uncertainties.

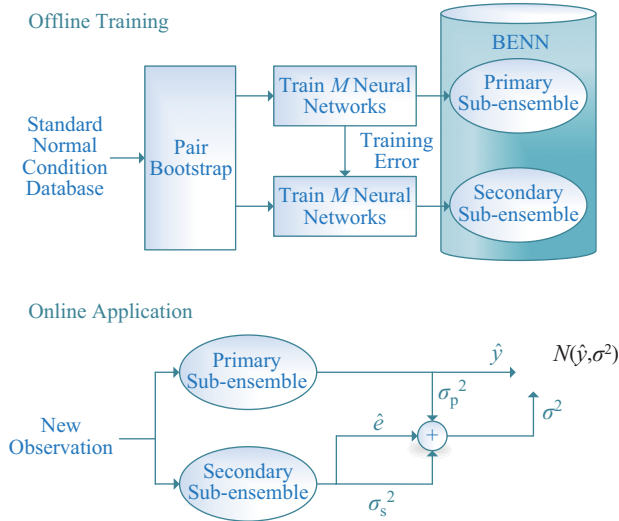


Fig. 2. Offline training and online application of BENN.

At the offline stage, the data to train each NN is sampled using the pair bootstrap method which is a general statistical inference process for regression analysis. Given a database $\Delta = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$, the pair bootstrap method

is to uniformly sample N observations with replacement. Those sampled observations constitute a bootstrapped dataset $\mathcal{D}_i = \{(\mathbf{x}'_1, y'_1), \dots, (\mathbf{x}'_N, y'_N)\}$ to train the i th NN. The pair bootstrap is an essential step to create the diversity among individual NNs. The primary and secondary sub-ensembles are trained sequentially, where the primary sub-ensemble is trained with the standard normal condition database extracted from historical SCADA observations while the secondary sub-ensemble is trained with the training error from primary sub-ensemble as the training target. The training error e'_i on a bootstrapped pair $(\mathbf{x}'_i, y'_i)$ is evaluated as follows [37],

$$e'_i = \left(y'_i - \frac{1}{M} \sum_{j=1}^M p_j(\mathbf{x}'_i) \right)^2 \quad (19)$$

where p_j represents the function mapped by the j th NN in the primary sub-ensemble, and M is the size of the sub-ensemble, i.e., the number of NN in the sub-ensemble.

At the online stage, for a new observation $(\mathbf{x}, y)$, the distribution in (2) is estimated as follows [37],

$$\hat{y} = \frac{1}{M} \sum_{j=1}^M p_j(\mathbf{x}) \quad (20)$$

$$\sigma^2 = \sigma_p^2(\mathbf{x}) + \hat{e} + \sigma_s^2(\mathbf{x}) \quad (21)$$

where

$$\sigma_p^2(\mathbf{x}) = \frac{1}{M-1} \sum_{j=1}^M (p_j(\mathbf{x}) - \hat{y})^2 \quad (22)$$

$$\hat{e} = \frac{1}{M} \sum_{j=1}^M s_j(\mathbf{x}) \quad (23)$$

$$\sigma_s^2(\mathbf{x}) = \frac{1}{M-1} \sum_{j=1}^M (s_j(\mathbf{x}) - \hat{e})^2 \quad (24)$$

where s_j represents the function mapped by the j th NN in the secondary sub-ensemble.

In (20)–(24), the distribution mean of target parameter, i.e. $\hat{y}$, is the output of the primary sub-ensemble. The associated uncertainty, i.e. σ^2 , is modelled as from three sources: the data, the primary sub-ensemble model, and the secondary sub-ensemble model. The model uncertainties, i.e. $\sigma_p^2(\mathbf{x})$ and $\sigma_s^2(\mathbf{x})$, are evaluated as the variance of the outputs from individual NNs, while the data uncertainty is evaluated by the secondary sub-ensemble.

V. CASE STUDY

The proposed approach is demonstrated on a SCADA dataset collected from a 1.5 MW type-3 onshore WT at Ningxia, China. This SCADA dataset was provided by GD POWER DEVELOPMENT CO., LTD. This section aims to verify the effectiveness of the proposed probabilistic approach and its advantages over the conventional deterministic approach.

A. Dataset Description

The selected dataset contains 5-month SCADA data between May 2017 and September 2017. Depending on the sensors placed in the WT, the SCADA system takes measurements of 79 parameters, covering the main WT components including gearbox, generator, blade, pitch control system, yaw system, hydraulic system, and electronic converter. Some parameters

are external ones to reflect the measurements on the environment condition, including wind speed, wind direction, and ambient temperature. Those external parameters are important and must be considered in WTCM, because their variations can lead to significant changes in internal parameters and induce uncertainties in parameter regression.

The SCADA system takes measurements every 10 minutes, which accumulates a total of 20,159 observations (the observations between 2nd August and 8th August are missing due to the unavailability of the SCADA system). Based on the method described in Section II-A-2) with $t_{\text{ahead}} = 4$ hours, 11,898 observations are selected as the normal condition observations. Among them, 80% are randomly selected as training normal observations to train and validate NBM and the remaining 20% are used as testing normal observations to test NBM.

B. Fault Event Selection

After scanning the dataset, two fault events are selected to test the performance of the proposed CM approaches. Fault 1 is an over-temperature fault indicated by gearbox oil. For this fault, the gearbox oil temperature is used as the target parameter to indicate the fault trend. According to measurements as shown in Fig. 3(a), it can be seen that the gearbox oil temperature remains at a high level in May and triggers the fault alarm on 13th May as it exceeds the 80°C over-temperature warning limit. After that, 9 more alarms were subsequently issued between 17th May and 3rd June, which indicated severe gearbox problems. The maintenance was then scheduled for 3rd June, which brought the gearbox oil temperature back to normal.

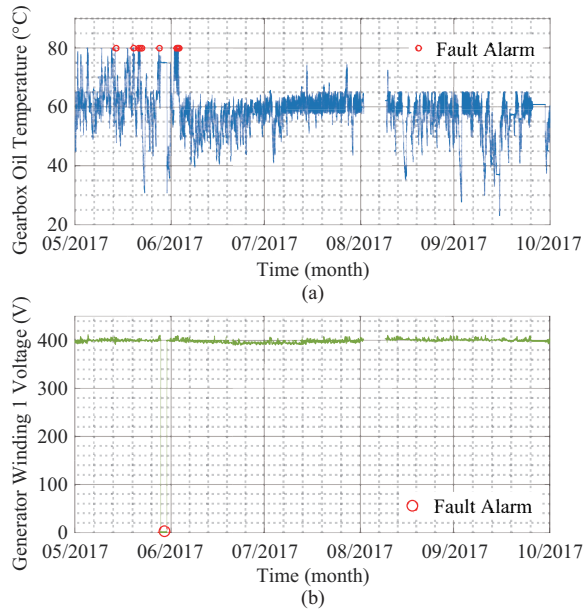


Fig. 3. The collected measurements on target parameters to indicate (a) fault 1 and (b) fault 2.

Fault 2 is a generator fault which occurred on 28th May. The target parameter for this fault is the voltage on generator winding 1. In Fig. 3(b), it can be seen that the generator winding voltage dropped sharply to zero. The problem persisted for more than 2 days and was fixed on 30th May.

C. Feature Selection

For each target parameter, the RELIEFF algorithm is performed on all the other parameters to select its critical regression features. Importance weight is the RELIEFF metric to select dependent parameters. The ranked importance weights generated by RELIEFF are shown in Fig. 4 where the features with higher importance weights should contain more information to describe the target parameters. In this study, each target parameter is modelled by the top 10 parameters in the rank, which are listed in Table I.

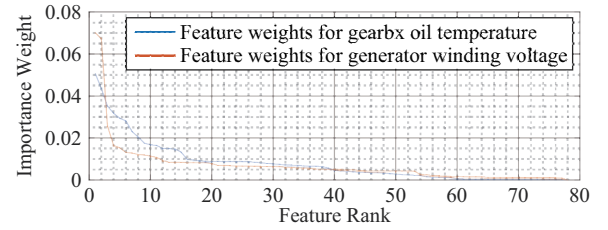


Fig. 4. The feature importance weights for each target parameter.

TABLE I
SELECTED DEPENDENT PARAMETERS

Target Parameters	Rank	Selected Dependent Parameters
Gearbox oil temperature	1	Gearbox inlet oil temperature
	2	Gearbox input shaft temperature
	3	Gearbox output shaft temperature
	4	Generator bearing temperature
	5	Yaw position
	6	Rotor-side main shaft bearing temperature
	7	Nacelle temperature
	8	Grid frequency
	9	Rotor position
	10	60 s average wind direction
Generator winding 1 voltage	1	Generator winding 2 voltage
	2	Generator winding 3 voltage
	3	Generator reactive power
	4	Yaw position
	5	Converter power
	6	Grid frequency
	7	Ambient temperature
	8	Rotor position
	9	Rotor-side main shaft bearing temperature
	10	Nacelle temperature

D. Normal Behavior Modelling

Two probabilistic NBMs, i.e. GP-based and BENN-based, are trained for each target parameter using the 1,1898 normal condition observations. The squared exponential function (11) is the covariance function of GP, and the hyperparameters are optimized by (12). For BENN, a fast learning algorithm called extreme learning machine (ELM) [38] is adopted to train individual NNs. ELM randomizes the input weights and biases of NN, so the traditional back-propagation training process is skipped, which significantly enhances the learning speed. The fast learning capability of ELM helps alleviate the high burden of training multiple NNs in an ensemble form.

The anomaly detection criterion on each target parameter is decided under a 5-fold cross validation structure. As in the validation results, the histogram of the two control variables (i.e. smoothed residual z_t and variance σ^2), the fitted gamma distribution, and the control limit η are shown in Fig. 5.

The smoothed residuals z_t here are the moving average of residual over the past 1 hour. The use of z_t and σ^2 as control variables respectively represents deterministic and probabilistic approaches. The gamma distributions are fitted using the maximum likelihood estimation method. It can be seen that the control variables follow the trend of their fitted distribution for all target parameters and adopted algorithms, which verifies the gamma distribution assumption on control variables. The control limits η of both control variables are calculated based on the confidence limits p following (7) and (8). Moreover, to generalize the performance of the WTCM, three p values (i.e. 99%, 99.5%, and 99.9%) are used in this case study,

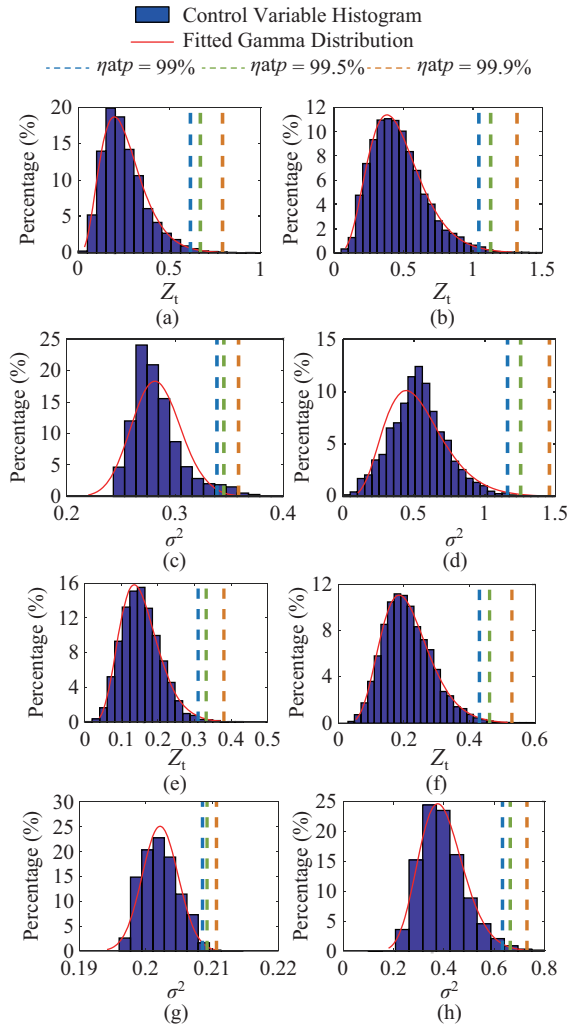


Fig. 5. The histogram of control variables, the fitted gamma distribution, and the control limits η for different scenarios: (a) gearbox oil temperature as the target parameter using GP algorithm with smoothed residual z_t as control variable; (b) gearbox oil temperature as the target parameter using BENN algorithm with smoothed residual z_t as control variable; (c) gearbox oil temperature as the target parameter using GP algorithm with variance σ^2 as control variable; (d) gearbox oil temperature as the target parameter using BENN algorithm with smooth variance σ^2 as control variable; (e) generator winding 1 voltage as the target parameter using GP algorithm with smoothed residual z_t as control variable; (f) generator winding 1 voltage as the target parameter using BENN algorithm with smoothed residual z_t as control variable; (g) generator winding 1 voltage as the target parameter using GP algorithm with variance σ^2 as control variable; and (h) generator winding 1 voltage as the target parameter using BENN algorithm with variance σ^2 as control variable.

which results in different control limits η . Using three p values here does not aim to optimize p for WTCM, but to test the sensitivity of p and η on CM and fault detection.

E. Fault Detection Accuracy

The NBMs and their control limits are applied on the testing datasets and consist of testing normal observations and the observations with SCADA fault alarms (indicated by the red circles in Fig. 3). The testing of normal observations represents the unseen normal samples while the observations with fault alarms serve as the faulty samples which should be accurately detected by NBMs.

In this case study, considering the anomaly detection nature of the proposed approach, four metrics are used to evaluate the classification performance of the NBMs: miss alarm rate (MAR), false alarm rate (FAR), accuracy (ACC), and Matthews correlation coefficient (MCC) [39],

$$MAR = fn / (fn + tp) \quad (25)$$

$$FAR = fp / (fp + tn) \quad (26)$$

$$ACC = (tp + tn) / (tp + tn + fp + fn) \quad (27)$$

$$MCC = \frac{tp \times tn - fp \times fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}} \quad (28)$$

where tp is the number of faulty samples that are classified as abnormal; tn is the number of normal samples that are classified as normal; fp is the number of normal samples that are classified as abnormal; and fn is the number of faulty samples that are classified as normal.

In the testing dataset for WT fault detection, data imbalance issue is frequently encountered since the fault alarms are rarely experienced in practice. The simple ACC metrics put equal weights on samples in different classes, which is inadequate and can lead to poor generalization in WT fault detection. In comparison, MCC is generally deemed as a balanced metric to evaluate the classification accuracy. The value of MCC is high only if the classifier can perform well on both normal and abnormal observations.

The fault detection accuracy on the testing dataset for different scenarios is listed in Table II. The following points can be summarized from Table II:

1) For WT faults with apparent indication, such as fault 2 with sudden voltage drop as shown in Fig. 3(b), they can be accurately detected under all scenarios. This can be observed from the 0.00% MAR.

2) For WT faults with a marginal detection boundary, such as fault 1 where its alarm is issued based on a thresholding strategy, the probabilistic approaches outperform the deterministic approaches in terms of fault detection accuracy (i.e. using σ^2 as the control variable achieves lower MAR, lower FAR, higher ACC, and higher MCC).

3) As a guide to select the best probabilistic algorithm, the GP algorithm outperforms the BENN algorithm in terms of fault detection accuracy.

4) As a probabilistic algorithm, GP shows 0.00% MAR on both faults, which promotes a higher p value to achieve lower FAR. In doing so, the overall fault detection accuracy of GP can be enhanced without generating false alarms on the faults.

5) The minimum ACC in Table II is 98.59% which can be deemed as being sufficiently accurate in WT fault detection, compared to ACC metrics that generates similar values for all scenarios. MCC metrics show significant variation, which can truly reflect the performance deviation between different algorithms, control variables, and control limits.

F. Fault Detection Earliness

The NBMs are applied on the unseen observations and their control charts before the physical alarms of the two faults are shown in Fig. 6. The alarms provided by SCADA systems are indicated as red points in the figure. For each fault, the results of GP and BENN are respectively presented in the left and right subfigures. In the control charts, the potential component faults are detected when the control variable (i.e. either z_t or σ^2) exceeds its control limit η . The time instances to detect the faults are labelled in each subfigure. The following points can be noted from Fig. 6:

1) The probabilistic approaches are much more efficient at

fault detection compared to the deterministic approaches for both faults and all the tested algorithms. Taking the subfigures (e) and (g) as examples, based on different p values, fault 1 is detected at 22.40 by the GP-based probabilistic approach, which are 9.5, 9.7 and 10 hours earlier than the deterministic approach (i.e. 08.10, 08.20, and 08.40 on the next day).

2) In the probabilistic approach, the GP algorithm outperforms the BENN algorithm in terms of the fault detection timing. To compare the subfigure (c) and (g) with subfigure (d) and (h), the GP algorithm detects the faults around 20 minutes earlier than the BENN algorithm.

3) The probabilistic approach is insensitive to the p value in terms of fault detection timing.

G. Sensitivity Test on t_{ahead}

According to the method described in Section II.A-2, the normal condition observations are decided by WT status code and t_{ahead} . The value of t_{ahead} should be carefully selected as it can affect the fault detection performance. In this paper,

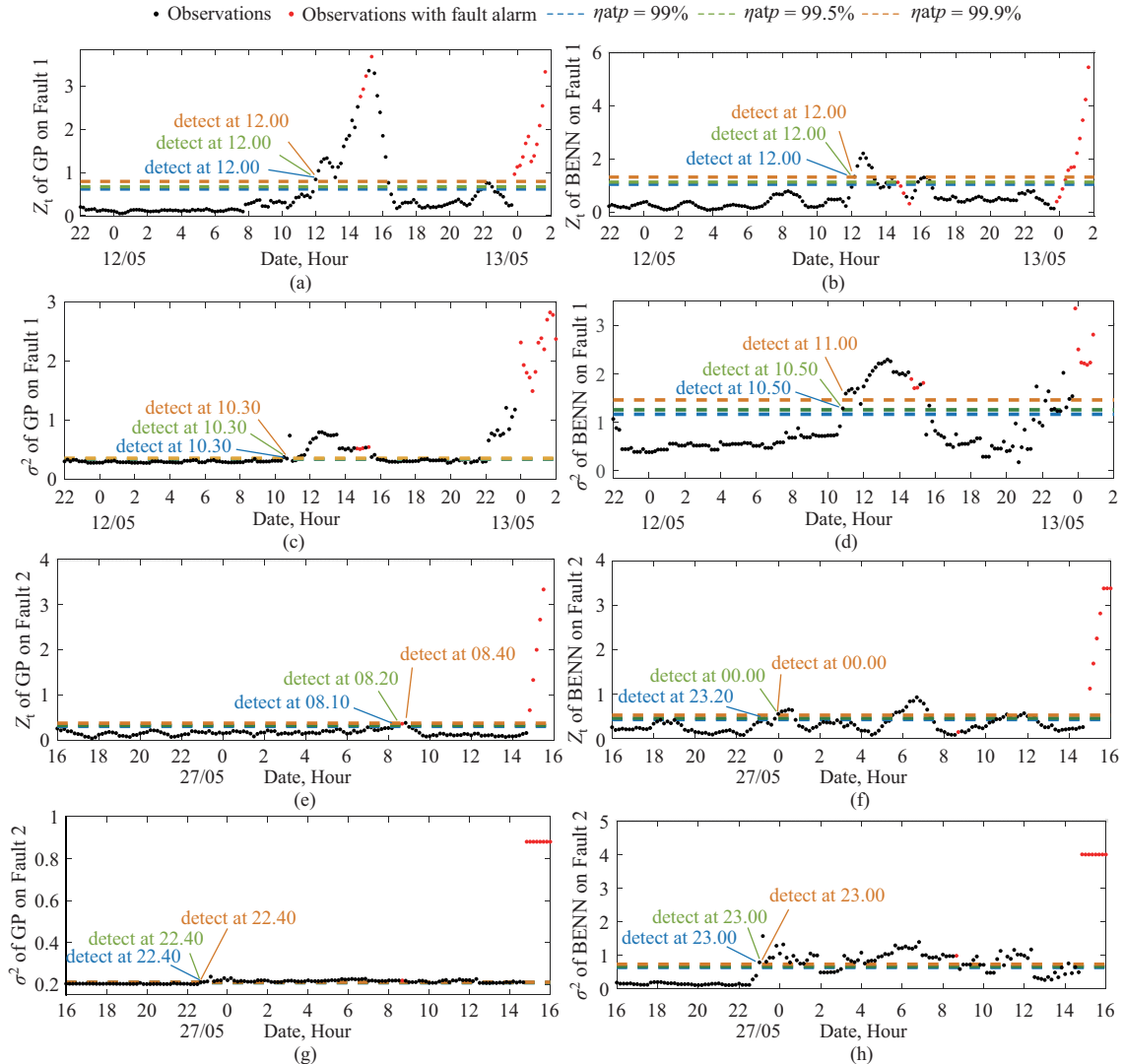


Fig. 6. The control charts of (a) smoothed GP residuals before fault 1 alarms; (b) smoothed BENN residuals before fault 1 alarms; (c) GP variances before fault 1 alarms; (d) BENN variances before fault 1 alarms; (e) smoothed GP residuals before fault 2 alarms; (f) smoothed BENN residuals before fault 2 alarms; (g) GP variances before fault 2 alarms; and (h) BENN variances before fault 2 alarms.

TABLE II
FAULT DETECTION ACCURACY

Fault/Algorithm/ Control Variable	$p(\%)$	MAR(%)	FAR(%)	ACC(%)	MCC(%)
Fault 1/GP/ z_t	99	16.00	0.71	98.67	82.89
	99.5	18.00	0.55	98.75	83.48
	99.9	23.00	0.13	98.95	85.59
Fault 1/GP/ σ^2	99	0.00	1.13	98.91	88.23
	99.5	0.00	0.88	99.15	90.51
	99.9	0.00	0.38	99.64	95.60
Fault 1/BENN/ z_t	99	27.00	0.46	98.47	78.88
	99.5	30.00	0.34	98.47	78.52
	99.9	31.00	0.17	98.59	80.11
Fault 1/BENN/ σ^2	99	14.00	0.46	98.99	86.80
	99.5	17.00	0.29	99.03	87.00
	99.9	23.00	0.04	99.03	86.74
Fault 2/GP/ z_t	99	0.00	0.84	99.25	96.33
	99.5	0.00	0.34	99.70	98.49
	99.9	0.00	0.04	99.96	99.81
Fault 2/GP/ σ^2	99	0.00	0.04	99.96	99.81
	99.5	0.00	0.00	100.00	100.00
	99.9	0.00	0.00	100.00	100.00
Fault 2/BENN/ z_t	99	0.00	1.13	98.99	95.13
	99.5	0.00	0.55	99.51	97.58
	99.9	0.00	0.25	99.78	98.86
Fault 2/BENN/ σ^2	99	0.00	0.97	99.14	95.81
	99.5	0.00	0.46	99.59	97.94
	99.9	0.00	0.21	99.81	99.05

a sensitivity test is carried out to investigate the effect of t_{ahead} and to select the best value. This sensitivity test is to screen the fault detection accuracy of NBM over a range of t_{ahead} values. Since the case study results have indicated that GP outperforms BENN in terms of fault detection accuracy and earliness, GP is used as the probabilistic algorithm in this sensitivity test. Moreover, in Table II, fault 1 results in larger accuracy variation over different testing scenarios compared to fault 2, so fault 1 is selected as the fault for this sensitivity test.

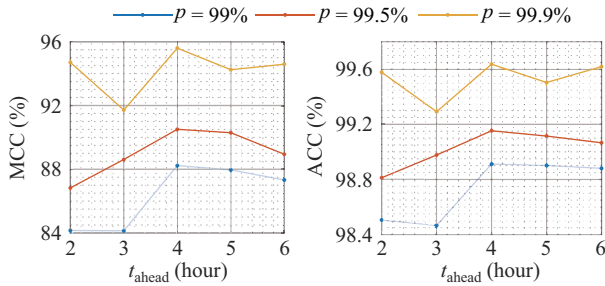


Fig. 7. Sensitivity test results on t_{ahead} in terms of (a) MCC and (b) ACC.

The testing results are shown in Fig. 7 where MCC and ACC are used as the accuracy metrics. The t_{ahead} varies between 2 and 6 hours. On the one hand, when t_{ahead} is small, the emerging fault observations tend to be recognized as normal, and the NBM trained with this abnormal data may provide inaccurate results. On the other hand, when t_{ahead} is large, a number of normal condition observations will not be learnt by NBM, which tends to generate false alarms on unseen observations. From Fig. 7, it can be seen that the MCC and ACC achieves its maximum at $t_{\text{ahead}} = 4$ for all p values, which verifies the optimal t_{ahead} parameter selection in this case study.

VI. CONCLUSION

The SCADA-based data-driven approaches have been identified as a cost-effective solution for WTCM and have shown its potential in practical applications. This paper proposes an AI-based probabilistic anomaly detection approach that can not only provide deterministic estimation on WT component condition but also evaluate the associated uncertainties. In the proposed approach, the abnormal conditions of the WT component are detected based on the evaluated uncertainties to provide a noise-free indication on emerging faults. Compared to the conventional deterministic approaches with residual-based anomaly detection criterion, the proposed probabilistic approach eliminates the noise filtering process, which tends to detect the faults at earlier. The proposed approach was tested on a SCADA dataset collected from an operational WT in China, and two probabilistic algorithms, GP and BENN, were evaluated in the test. The results demonstrated that the proposed probabilistic approach demonstrated higher accuracy and earlier fault detection compared to the conventional deterministic approach. The GP algorithm outperformed the BENN algorithm in terms of its higher accuracy and earlier fault detection.

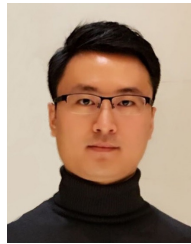
REFERENCES

- [1] D. E. Bently and C. T. Hatch, *Fundamentals of Rotating Machinery Diagnostics*, New York: ASME Press, 2003.
- [2] A. Kusiak and W. Y. Li, "The prediction and diagnosis of wind turbine faults," *Renewable Energy*, vol. 36, no. 1, pp. 16–23, Jan. 2011.
- [3] W. Qiao and D. G. Lu, "A survey on wind turbine condition monitoring and fault diagnosis - Part I: components and subsystems," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6536–6545, Oct. 2015.
- [4] G. Helbing and M. Ritter, "Deep learning for fault detection in wind turbines," *Renewable and Sustainable Energy Reviews*, vol. 98, pp. 189–198, Dec. 2018.
- [5] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 636–653, May 2010.
- [6] P. F. Odgaard and J. Stoustrup, "A benchmark evaluation of fault tolerant wind turbine control concepts," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 3, pp. 1221–1228, May 2015.
- [7] S. Dey, P. Pisu, and B. Ayalew, "A comparative study of three fault diagnosis schemes for wind turbines," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 5, pp. 1853–1868, Sep. 2015.
- [8] W. Yang, P. J. Tavner, and M. R. Wilkinson, "Condition monitoring and fault diagnosis of a wind turbine synchronous generator drive train," *IET Renewable Power Generation*, vol. 3, no. 1, pp. 1–11, Mar. 2009.
- [9] Q. Huang, D. Jiang, L. Hong, and Y. Ding, "Application of wavelet neural networks on vibration fault diagnosis for wind turbine gearbox," in *Proceedings 5th International Symposium on Neural Networks*, 2008, pp. 313–320.
- [10] J. J. Christensen, C. Andersson and S. Gutt, "Remote condition monitoring of vestas turbines", Technical Track – Operation & Maintenance, in *Proc. EWEK 2009*, Marseille, France, Mar. 16–19, 2009.
- [11] P. Caselitz and J. Giebhardt, "Rotor condition monitoring for improved operational safety of offshore wind energy converters," *Journal of Solar Energy Engineering*, vol. 127, no. 2, pp. 253–261, Sep. 2005.
- [12] B. F. Sørensen, L. Lading, P. Sendrup, M. McGugan, C. P. Debel, O. J. D. Kristensen, G. C. Larsen, A. M. Hansen, J. Rheinländer, J. Rusborg, and J. D. Vestergaard, "Fundamentals for remote structural health monitoring of wind turbine blades—A preproject," Risø National Laboratory, Forskningscenter Risø, Roskilde, Denmark. Risø-R, No. 1336 (EN), May 2002.
- [13] Y. Qu, E. Bechhoefer, D. He, and J. Zhu, "A new acoustic emission sensor based gear fault detection approach," *International Journal of Prognostics and Health Management*, vol. 4, pp. 1–14, Jul. 2013.
- [14] R. Y. Li, S. U. Seckiner, D. He, E. Bechhoefer, and P. Menon, "Gear fault location detection for split torque gearbox using AE sensors," *IEEE*

- Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1308–1317, Nov. 2012.
- [15] A. Widodo E. Y. Kim, J. D. Son, B. S. Yang, A. C. C. Tan, D. S. Gu, B. K. Choi, and J. Mathew, "Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine," *Expert Systems with Applications*, vol. 36, no. 3, pp. 7252–7261, Apr. 2009.
 - [16] O. Bennouna, N. Héraud, H. Camblong, and M. Rodriguez, "Diagnosis of the doubly-fed induction generator of a wind turbine," *Wind Engineering*, vol. 29, no. 5, pp. 431–447, Sep. 2005.
 - [17] X. Gong and W. Qiao, "Imbalance fault detection of direct-drive wind turbines using generator current signals," *IEEE Transactions on Energy Conversion*, vol. 27, no. 2, pp. 468–476, Jun. 2012.
 - [18] X. Gong and W. Qiao, "Current-based mechanical fault detection for direct-drive wind turbines via synchronous sampling and impulse detection," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 3, pp. 1693–1702, Mar. 2015.
 - [19] M. R. Shahriar, P. Borghesani, G. Ledwich, and A. C. C. Tan, "Performance analysis of electrical signature analysis-based diagnostics using an electromechanical model of wind turbine," *Renewable Energy*, vol. 116, pp. 15–41, Feb. 2018.
 - [20] W. Qiao and D. G. Lu, "A survey on wind turbine condition monitoring and fault diagnosis - Part II: signals and signal processing methods," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6546–6557, Oct. 2015.
 - [21] A. Kusiak and A. Verma, "A data-mining approach to monitoring wind turbines," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 150–157, Jan. 2012.
 - [22] A. Kusiak and A. Verma, "A data-driven approach for monitoring blade pitch faults in wind turbines," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 1, pp. 87–96, Jan. 2011.
 - [23] W. X. Yang, R. Court, and J. S. Jiang, "Wind turbine condition monitoring by the approach of SCADA data analysis," *Renewable Energy*, vol. 53, pp. 365–376, May 2013.
 - [24] P. Guo, D. Infield, and X. Y. Yang, "Wind turbine generator condition-monitoring using temperature trend analysis," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 124–133, Jan. 2012.
 - [25] P. B. Dao, W. J. Staszewski, T. Barszcz, and T. Uhl, "Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data," *Renewable Energy*, vol. 116, pp. 107–122, Feb. 2018.
 - [26] Z. Zaher, S. D. J. McArthur, D. G. Infield, and Y. Patel, "Online wind turbine Fault detection through automated SCADA data analysis," *Wind Energy*, vol. 12, no. 6, pp. 574–593, Sep. 2009.
 - [27] A. Kusiak and A. Verma, "Analyzing bearing faults in wind turbines: a data-mining approach," *Renewable Energy*, vol. 48, pp. 110–116, Dec. 2012.
 - [28] P. Bangalore and L. B. Tjernberg, "An artificial neural network approach for early fault detection of gearbox bearings," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 980–987, Mar. 2015.
 - [29] H. S. Zhao, H. H. Liu, W. J. Hu, and X. H. Yan, "Anomaly detection and fault analysis of wind turbine components based on deep learning network," *Renewable Energy*, vol. 127, pp. 825–834, Nov. 2018.
 - [30] M. Schlechtingen, I. F. Santos, and S. Achiche, "Wind turbine condition monitoring based on SCADA data using normal behavior models. Part I: system description," *Applied Soft Computing*, vol. 13, no. 1, pp. 259–270, Jan. 2013.
 - [31] W. G. Yang, C. Liu, and D. X. Jiang, "An unsupervised spatiotemporal graphical modeling approach for wind turbine condition monitoring," *Renewable Energy*, vol. 127, pp. 230–241, Nov. 2018.
 - [32] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proceedings of the 9th International Workshop on Machine Learning*, 1992, pp. 249–256.
 - [33] M. Robnik-Šikonja, and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, no. 1–2, pp. 23–69, 2003.
 - [34] C. K. I. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Advances in Neural Information Processing Systems 8*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds. Cambridge, MA: MIT Press, 1996, pp. 514–520.
 - [35] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Cambridge, MA: MIT Press, 2006.
 - [36] Y. Ren, L. Zhang, and P. N. Suganthan, "Ensemble classification and regression-recent developments, applications and future directions," *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 41–53, Feb. 2016.
 - [37] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1033–1044, May 2014.
 - [38] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.
 - [39] S. Boughorbel, F. Jarray, and M. El-Anbari, "Optimal classifier for imbalanced data using Matthews correlation coefficient metric," *PLoS One*, vol. 12, no. 6, pp. e0177678, Jun. 2017.



Yuchen Zhang (S'15–M'18) received the B.E., B.Com. and Ph.D. degrees from the University of New South Wales, Sydney, Australia, in 2013, 2013, and 2018, respectively. He is currently a Research Associate at the University of New South Wales, Sydney, Australia. His research interests include power system stability and control, wind farm condition monitoring and fault diagnostics, data analytics, and machine learning applications in power engineering.



Meng Li received the B.Sc. degree from the China University of Mining and Technology, Xuzhou, China in 2011, the M.E. degree from the Anhui University of Science & Technology, Huainan, China in 2013. He is currently a Ph.D. candidate with China University of Mining and Technology. His research interest covers sensor and sensing technology, fault diagnosis and Terahertz Technology applications.



Zhao Yang Dong (M'99–SM'06–F'17) obtained his Ph.D. degree from the University of Sydney, Australia in 1999. He is currently the SHARP Professor with the University of New South Wales, Australia, Director of UNSW Digital Grid Future Institute, Director of ARC Research Hub on Integrated Energy Storage Solutions, and Vice Chairman of Global Chinese Electrical and Energy Technology Industry Alliance. He was previously Professor and Head of School of Electrical and Information Engineering, University of Sydney, and Ausgrid Chair and Director of the Ausgrid Centre for Intelligent Electricity Networks, the University of Newcastle, Australia. He also held industrial positions with Transend Networks (now TAS Networks), Australia. His research interests include smart grid, power system planning, power system security, renewable energy systems, electricity market, load modelling, and computational intelligence and its application in power engineering. He is an editor of *IEEE Transactions on Smart Grid*, *IEEE Power Engineering Letters* and *IET Renewable Power Generation*.



Ke Meng (M'10) received the Ph.D. degree in Electrical Engineering from the University of Queensland, Brisbane, QLD, Australia in 2009. He is currently a Senior Lecturer with the School of Electrical Engineering and Telecommunications, the University of New South Wales, Sydney, NSW, Australia. He was previously a Lecturer with the School of Electrical and Information Engineering, the University of Sydney, Sydney, Australia. His research interests include pattern recognition, power system stability analysis, wind power and energy storage.